

APPLICATION OF LINEAR ALGEBRA IN GENETICS

➤ **Why Should One Care About Linear Algebra?**

Linear Algebra is an indispensable part of our lives. Whether it's using Google Maps to navigate an unfamiliar place, receiving recommendations on OTT platforms like Netflix, understanding traffic flow, managing business and finance, or its wide-ranging applications in AI, such as facial recognition—linear algebra is everywhere.

One particularly significant application from a biological standpoint is in genetics.

In this write-up, we'll explore how linear algebra is applied to determine the genotype distribution of a certain trait in a population. By modelling genetic variation using matrices, researchers can analyze data more efficiently and accurately. This can lead to insights about heredity, evolution, and even the likelihood of certain traits manifesting in future generations.

Understanding these concepts not only enhances our knowledge of biology but also illustrates the power of linear algebra in solving real-world problems. From predicting outcomes in genetics to optimizing algorithms in technology, the applications are vast and impactful.

➤ **Genetics, Traits, and Genotypes**

Genetics is a branch of biology that focuses on the study of genes, genetic variation, and heredity in organisms. Throughout the study of genetics, various factors contribute to determining the characteristics of an organism. Through natural selection and adaptation, the genes of an organism change to better fit their surroundings, leading to the loss of certain genetically determined characteristics, or traits, while dominant traits persist throughout their existence.

Genotype refers to the genetic makeup of an organism. Broadly speaking, we can examine the genotype distribution of a particular trait in a population. For instance, consider a heritable trait governed by a single gene on a chromosome. Understanding the distribution of this genotype helps researchers predict how traits may manifest in future generations and how they influence the overall diversity within a population.

This intersection of genetics and linear algebra allows for a more nuanced understanding of biological processes and the inheritance of traits.

➤ Mathematics Involved:-

In this write-up, we'll incorporate the following mathematical tools:

- Eigenvalues and Eigenvectors
- Matrix Representation
- Diagonalization of Matrices
- Inverse of Matrices

We will develop a model to depict the distribution of a genotype of a certain trait in a population. Let's understand this through a case study.

Suppose a farmer wants to maximize profit by catering to consumer preferences and observes that people prefer green grapes over black grapes. In this scenario, there are two types of alleles:

- G : Dominant allele (green grapes)
- g : Recessive allele (black grapes)

Using these alleles, we can model the genotype frequencies in the population. By applying concepts from linear algebra, such as matrix representation and eigenvalues, we can analyse the genotype distribution and predict how it might change over generations based on consumer demand. This mathematical approach not only helps the farmer make informed decisions but also illustrates the broader implications of genetics in agriculture and economics.

➤ Possible Genotypes:-

In the context of our study, there are three possible genotypes for grapes:

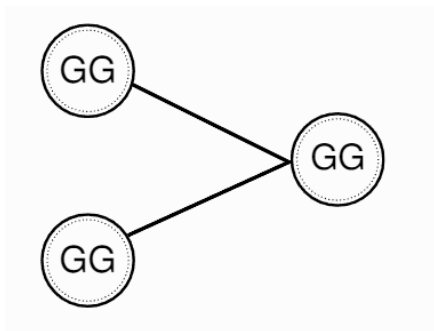
- 1). GG : Green grapes (homozygous dominant)
- 2). Gg : Green grapes (heterozygous)
- 3). gg : Black grapes (homozygous recessive)

➤ Possible Genotype Distribution

The farmer wants to initiate a program where each grape in the population is fertilized with a **GG** grape and subsequently replaced by one of its offspring. We need to compute an expression for its genotype distribution.

➤ Case (A): Fertilizing GG with GG

When two GG grapes are fertilized, the possible offspring are :-



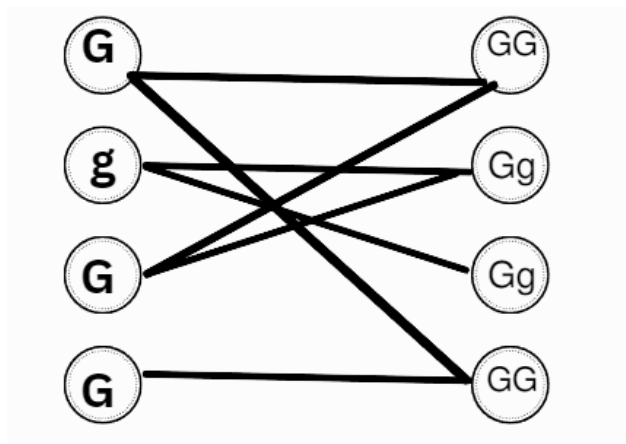
The distribution remains:

- GG: 100% (All offspring will be green grapes)

➤ Case (B): Fertilizing Gg with GG

When a “Gg” grape is fertilized with a “GG” grape, the possible offspring are:

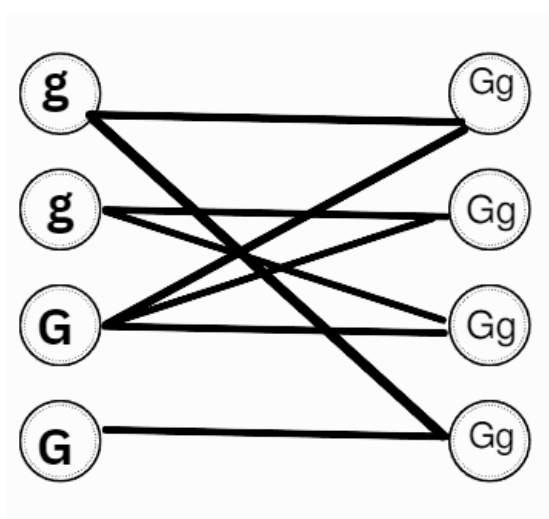
- “G” (from GG) + “G” (from Gg) = “GG”
- “g” (from Gg) + “G” (from GG) = “Gg”



The distribution will be:

- GG : 50%
- Gg : 50%

Case (c)
fertilizing " gg " with " GG ".



Summary of Genotype Distribution:-

By applying these fertilization scenarios, we can analyze how the genotype distribution of grapes changes over time. The farmer can use this model to estimate the expected proportions of each genotype in future generations, aiding in decisions about cultivation and market strategy. Using linear algebra techniques, such as matrix representation, can further refine these predictions and enhance the farmer's understanding of genetic dynamics in the population.

Probabilities are :

Genotype of off -spring	Case 1	Case 2	Case 3
GG	1	$\frac{1}{2}$	0
Gg	0	$\frac{1}{2}$	1
gg	0	0	0

Thus , from above table we get :

$$\text{Matrix (A) } = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Assuming the initial population of grapes consists of an equal number of each genotype, we have:

$$X_0 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

The distribution of the first generation of offspring can be represented as:

$$X_0 \rightarrow X_1 = A X_0 \longrightarrow (1)$$

Similarly, we can compute the distribution of future generations. For the nth generation, based on the given initial distribution, we can express it as:

$$X_n = A X_{(n-1)}$$

$$X_n = A X_{(n-2)}$$

$$X_n = A X_{(n-3)}$$

$$X_n = A X_{(n-4)}$$

.

.

.

$$X_n = A^n X_0$$

First we will find the eigenvalues and eigen vectors for the given matrix "A" :

Let ,

$\lambda \rightarrow \text{parameter}$ & $I \rightarrow \text{is identity matrix}$

$$A - \lambda I = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A - \lambda I = \begin{bmatrix} 1 - \lambda & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} - \lambda & 1 \\ 0 & 0 & -\lambda \end{bmatrix} \quad (\text{equation 2})$$

Now , $\det (A - \lambda) = 0$

$$|A - \lambda I| = (1 - \lambda)((1/2 - \lambda)(-\lambda)) - 1/2(0 - 0) + 0(\dots\dots) = 0$$

$$(1 - \lambda)(\lambda - (1/2)) \times (\lambda) = 0$$

$$\lambda = 0, 1/2, 1, \dots\dots$$

THEREFORE ,

The eigenvalues corresponding to matrix A are as follows ;

$$\lambda_1 = 1$$

$$\lambda_2 = 1/2$$

$$\lambda_3 = 0$$

➤ The diagonalisation of A involves a diagonal matrix “ D ” which is given by ;

$$D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- Now for finding the corresponding eigenvectors we need to use the equation:

$$Ax = \lambda x$$

Where,

A = matrix (A)

X = eigen-vector

λ = eigen-value

$$(A - \lambda I) x = 0$$

- **Case-1:**

Where,

$$\lambda_1 = 1$$

$$(A - \lambda_1 I(x_1)) = 0$$

$$D = \begin{bmatrix} 1 - 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} - 1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (\text{By Equation 2})$$

$$D = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} = 0$$

Therefore,

$$a_1 = 1, b_1 = 0, c_1 = 0$$

The Eigenvector V_1 Corresponding to λ_1 is $(1, 0, 0)^T$

- **Case 2:**

Where,

$$\lambda_2 = 1/2$$

$$(A - \lambda_2 I(x_2)) = 0$$

$$D = \begin{bmatrix} 1 - 1/2 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} - \frac{1}{2} & 1 \\ 0 & 0 & -1/2 \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 1/2 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1/2 \end{bmatrix} \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix} = 0$$

Therefore,

$$c_2 = 0 \text{ and } \frac{a_2}{2} + \frac{b_2}{2} = 0$$

$$a_2 = -b_2 = \mu$$

$$X_2 = \mu \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

The Eigenvector V_2 Corresponding to λ_2 is $(1, -1, 0)^T$

- **Case 3:**

Where,

$$\lambda_3 = 0$$

$$(A - \lambda_3 I(x_3)) = 0$$

$$D = \begin{bmatrix} 1 - 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} - 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_3 \\ b_3 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$D = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_3 \\ b_3 \\ c_3 \end{bmatrix} = 0$$

Therefore,

c_3 is Free Variable

$$c_3 = \mu, a_3 = \mu, b_3 = -2\mu$$

$$x_3 = \mu \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

The Eigenvector V_3 Corresponding to λ_3 is $(1, -2, 1)^T$

Now, the matrix A can be written in the form of diagonal matrix "D" and a matrix "P" as follows :

$$A = PDP^{-1}$$

Where,

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$P = [V_1 \quad V_2 \quad V_3]$$

$$= \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

The reason why we are decomposing the Matrix A in This way is too easy the process of computing the distribution. Otherwise, we would require to multiply matrix A, say 'n' times, which is a cumbersome task.

$$x_n = x_o A^n$$

$$A = PDP^{-1}$$

$$A^n = (PDP^{-1})^n$$

$$A^n = (PDP^{-1}PDP^{-1} \dots \dots \dots PDP^{-1}) \rightarrow n - times$$

$$A^n = (PDDDD \dots \dots \dots DP^{-1}) \rightarrow n - times$$

$$A^n = PD^nP^{-1}$$

$$D^n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}^n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \left(\frac{1}{2}\right)^n & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$p = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

- thus we apply Gauss-jordan method to compute the matrix inverse of “ p “

$$(P|I) = (I|P^{-1})$$

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \xrightarrow{R_2 \rightarrow (-1)R_2} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \xrightarrow{R_2 \rightarrow R_2 + (-2)R_3} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & -2 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right]$$

$$\xrightarrow{R_1 \rightarrow R_1 - R_2} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 & -2 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right] \xrightarrow{R_1 \rightarrow R_1 - R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & -1 & -2 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array} \right]$$

$$P^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A^n = PD^nP^{-1}$$

$$A^n = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \left(\frac{1}{2}\right)^n & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

Thus ,

$$X_n = A^n X_0$$

$$X_n = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \left(\frac{1}{2}\right)^n & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

Note: -

In this case study, we have assumed an equal number of each genotype in the initial population, i.e.,

$$X_0 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^T$$

However, the distribution does not always have to be equal and can be generalized as

$$(X_0 = (\alpha, \beta, \gamma)^T, \text{ where } (\alpha + \beta + \gamma = 1) \text{ and } (0 \leq \alpha, \beta, \gamma \leq 1)).$$

Continuing the calculation, upon expanding, we obtain:

$$X_n = \begin{bmatrix} 1 - \left(\frac{1}{2}\right)^n \\ \left(\frac{1}{2}\right)^n \\ 0 \end{bmatrix}$$

The ratio of genotype distribution is as follows :

$$GG : Gg : gg = \left(1 - \left(\frac{1}{2}\right)^n\right) : \left(\frac{1}{2}\right)^n : 0$$

Hence, the distribution can be computed as shown above. In the process, various mathematical tools associated with Linear Algebra are used, such as:

- Eigenvectors, Eigenvalues & Diagonalization
- Inverse of a Matrix
- Gauss-Jordan Method

REFERENCE :

- 1) <https://www.slideshare.net/slideshow/linear-algebra-to-solve-autosomal-inheritance/80315470>
- 2) <https://sites.math.washington.edu/~king/coursedir/m308a01/Projects/m308a01-pdf/kirkham.pdf>